BANA 4090: Chapter 4: ARIMA models (Part II)

Zhaohu(Jonathan) Fan

07/06/2021

- Main topics:
    - ARIMA models
        - A toy example
        - Stationarity and differencing

Prerequisites

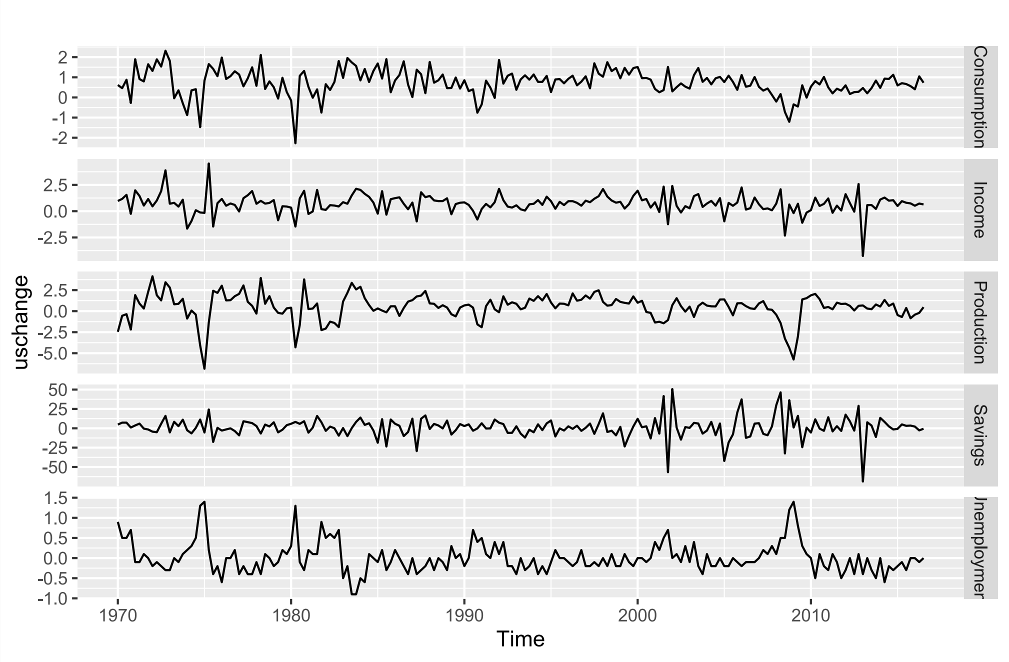
```r
# List of required (CRAN) packages
pkgs <- c(
  "ggplot2",# for drawing nicer graphics
  "fpp2",   # for using four simple forecasting models
  "forecast", #for using checkresiduals() function: a test of autocorrelation of the 
)

# Install required (CRAN) packages
for (pkg in pkgs) {
  if (!(pkg %in% installed.packages()[, "Package"])) {
    install.packages(pkg)
  }
}
```

A toy example

US personal consumption

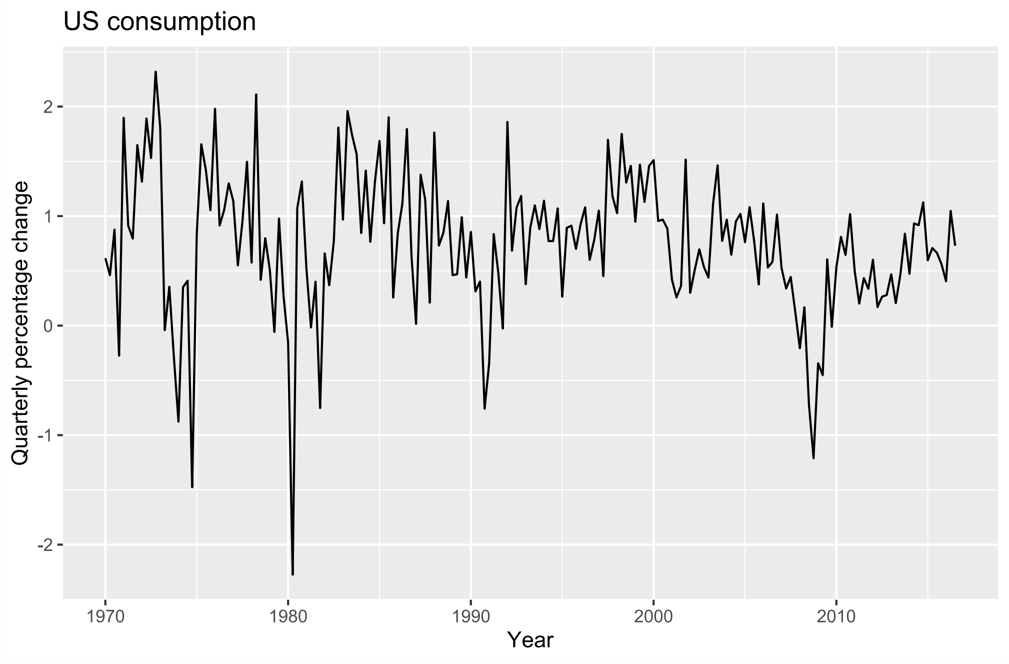
- Percentage changes in quarterly **personal consumption expenditure** , **personal disposable income** , **trended and seasonal** production, **savings** and **the unemployment rate** for the US, 1970 to 2016.
- Growth rates of personal consumption and personal income in the USA.
- Time series object of class `ts`.

```
autoplot(uschange, facet=TRUE)
```



US personal consumption

```r
autoplot(uschange[,"Consumption"]) +
  xlab("Year") +
  ylab("Quarterly percentage change") +
  ggtitle("US consumption")
```



US personal consumption

- ARIMA model

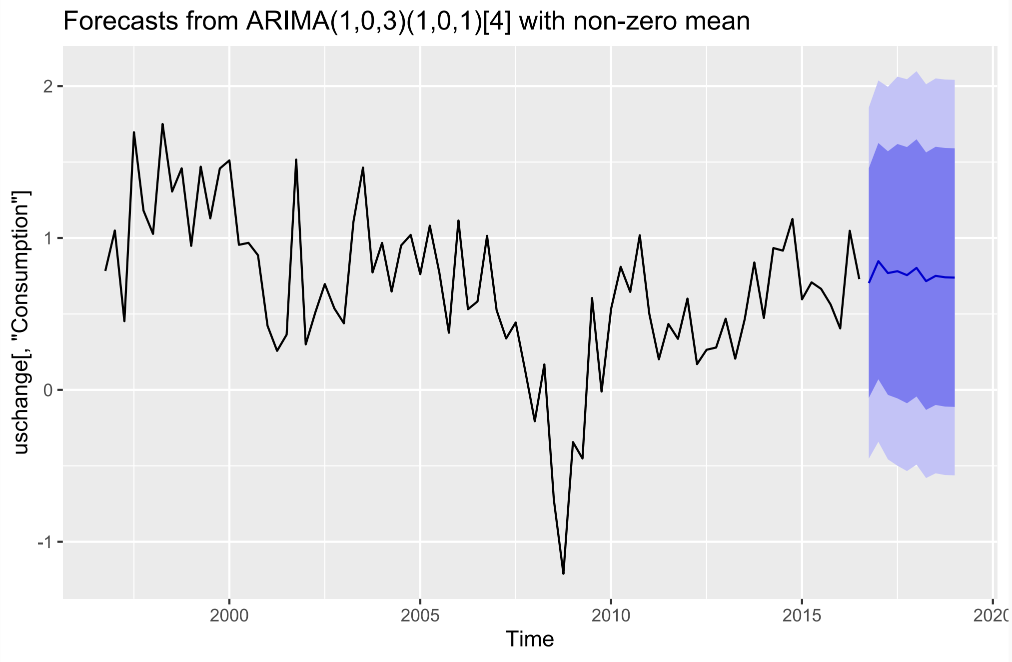
```
(fit <- auto.arima(uschange[,"Consumption"]))
```

```
## Series: uschange[, "Consumption"]
## ARIMA(1,0,3)(1,0,1)[4] with non-zero mean
##
## Coefficients:
##           ar1     ma1     ma2     ma3     sar1    sma1    mean
##       -0.3548  0.5958  0.3437  0.4111  -0.1376  0.3834  0.7460
## s.e.   0.1592  0.1496  0.0960  0.0825   0.2117  0.1780  0.0886
##
## sigma^2 estimated as 0.3481:  log likelihood=-163.34
## AIC=342.67   AICc=343.48   BIC=368.52
```

US personal consumption

- Forecast the US personal consumption for next 10 quarters.
- 80% prediction intervals.
- `%>%` pipe operator (function) in R means that you pass an intermediate result onto the next function.

```
fit %>% forecast(h=10) %>% autoplot(include=80)
```



Stationarity and differencing

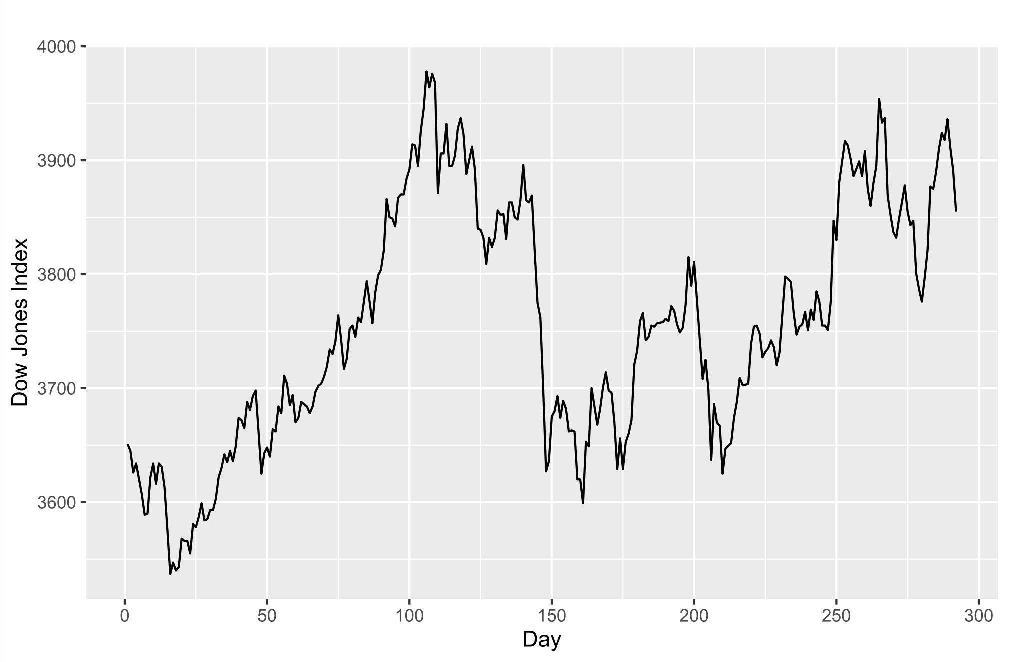
Stationarity

- A **stationary series** is:
    - roughly horizontal
    - constant variance
    - no patterns predictable in the long-term
- **Definition**
    - If $\{y_t\}$ is a stationary time series, then for all s, the distribution of $(y_t, \ldots, y_{t+s})$ does not depend on t.

Stationary?

- Dow-Jones index on 251 trading days ending 26 Aug 1994.

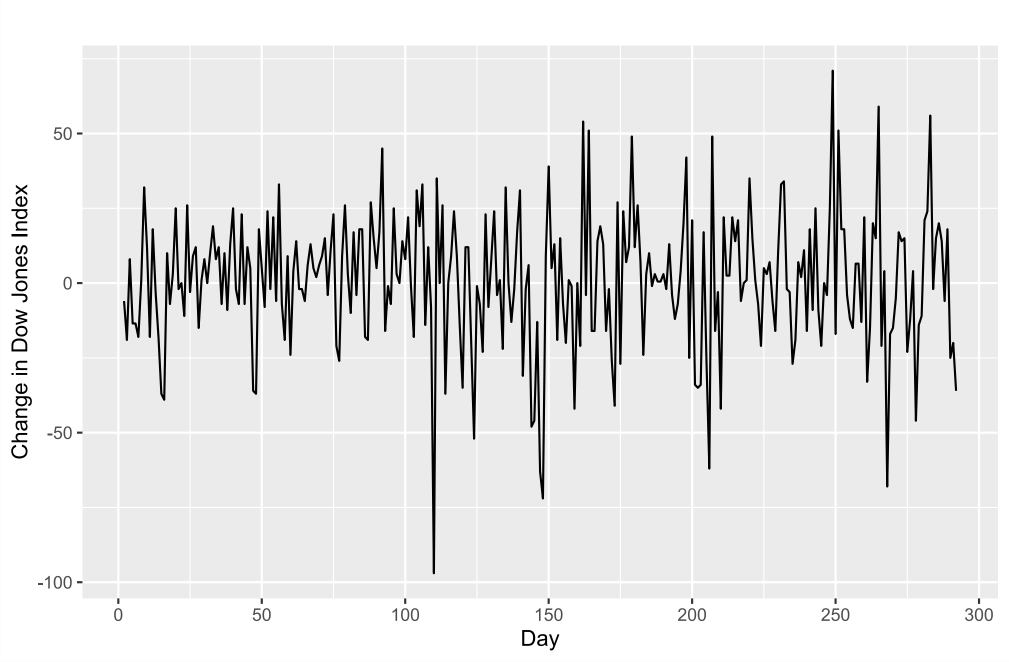
```
autoplot(dj) + ylab("Dow Jones Index") + xlab("Day")
```



Stationary?

- First-order differencing
- The differenced series is the **change** between each observation in the original series:
  - $y_t^{'} = y_t - y_{t-1}$. In R, `diff(data, lag=1)` or `diff(data)`.

```
autoplot(diff(dj)) + ylab("Change in Dow Jones Index") + xlab("Day")
```



Differencing

- Differencing helps to **stabilize the mean**.
- Second-order differencing
  - Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time.
- Seasonal differencing
  - A seasonal difference is the difference between an observation and the corresponding observation from the previous year.
  - $y_t' = y_t - y_{t-m}$, where m= number of seasons.
  - For monthly data m=12. In R, `diff(lag=12)`.
  - For quarterly data m=4. In R, `diff(lag=4)`.
- Note: The differenced series will have only $T-1$ values since it is not possible to calculate a difference y_1' for the first observation.

Interpretation of differencing

When both seasonal and first differences are applied…

- it makes no difference which is done first---the result will be the same.
- If **seasonality** is strong, we recommend that **seasonal differencing** be done first because sometimes the resulting series will be stationary and there will be no need for further first difference.

It is important that if differencing is used, the differences are interpretable.

Interpretation of differencing (cont'd)

- First differences are the change between **one observation and the next**;
- Seasonal differences are the change between **one year to the next**.
- But taking lag 3 differences for yearly data, for example, results in a model which cannot be sensibly interpreted.

Differencing

- Stationarity
  - Transformations help to **stabilize the variance**.
  - For ARIMA modelling, we need to **stabilize the mean**.

Augmented Dickey-Fuller Test (ADF)

- In general, the Augmented Dickey-Fuller test is defined as:
  - H_0 (null hypothesis): the data series is not stationary
  - H_a (alternative hypothesis): the data series is stationary.

- The `adf.test()` from the `tseries` package will do a Augmented Dickey-Fuller test
- Use `?adf.test` to read about this function.

```
library(tseries)
dj%>%adf.test
```

```
##
##      Augmented Dickey-Fuller Test
##
## data:  .
## Dickey-Fuller = -1.9872, Lag order = 6, p-value = 0.5816
## alternative hypothesis: stationary
```

- A p-value of ADF is greater than 0.05. Hence, we fail to reject the null hypothesis for this test, and we conclude that the data series is NOT stationary.

Augmented Dickey-Fuller Test (ADF)

```
diff((dj))%>%adf.test
```

```
##
##      Augmented Dickey-Fuller Test
##
## data:  .
## Dickey-Fuller = -6.7623, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

- A p-value of ADF is less than 0.05. Hence, we reject the null hypothesis for this test, and we conclude that the differenced data is stationary.